

AMENDMENTS TO THE CLAIMS

Claims 1, 3-13, 16, 19, 21-31, 34-36, 38-39, 41-48, and 51 amended.

All pending claims are reproduced below.

1. (Currently Amended) A method for selecting transformation rules for application to unstructured text content in customer accounts, comprising:

storing a plurality of customer accounts, each customer account comprising:

a structure content record of financial and personal information associated with a customer;

unstructured text content derived from an interaction with the customer;

and

an actual outcome of an event related to the customer;

providing a set of source tokens from the unstructured text content of the customer accounts, each source token associated with at least one of the structured content records, ~~each structured content record including an actual outcome;~~

applying candidate transformation rules to a set of source tokens to selectively produce tokens in response to the transformation rules;

determining for each candidate transformation rule a statistical measure of accuracy of ~~the a~~ predictive model for predicting outcomes of events related to the customers based on the actual outcomes of events in the customer accounts associated with the produced tokens; and

selecting transformation rules that ~~are likely to~~ improve the measure of accuracy of the predictive model.

2. (Original) The method of claim 1, further comprising:

associating each token produced by a transformation rule from a source token with structured content records associated with a source token.

3. (Currently Amended) The method of claim 1, wherein determining for each candidate transformation rule a statistical measure of accuracy comprises:

determining a number of correct and incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule.

4. (Currently Amended) The method of claim 1, wherein determining for each candidate transformation rule a statistical measure of accuracy comprises:

determining a distribution of correct and incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule.

5. (Currently Amended) The method of claim 1, wherein selecting transformation rules that ~~are likely to~~ improve a the measure of accuracy of the predictive model comprises:

selecting transformation rules that maximize the measure of accuracy of the predictive model.

6. (Currently Amended) The method of claim 1, wherein determining for a candidate transformation rule a statistical measure of accuracy of the predictive model comprises:

determining a number of correct predicted outcomes from the structured content records associated with a token produced by the transformation rule;

determining a number of correct predicted outcomes from the structured content records not associated with the produced token;

determining a number of incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule; and

determining a number of incorrect predicted outcomes from the structured content records not associated with the produced token.

7. (Currently Amended) The method of claim 1, wherein determining for each candidate transformation rule a statistical measure of accuracy of the predictive model comprises:

determining an information gain resulting from transformation rule.

8. (Currently Amended) The method of claim 1, wherein determining for each candidate transformation rule a statistical measure of accuracy of the predictive model comprises:

determining an Odds ratio for correct predicted outcomes in structured content records associated with a token produced by the transformation rule.

9. (Currently Amended) The method of claim 1, wherein determining for each candidate transformation rule a statistical measure of accuracy of the predictive model comprises:

determining a Chi-square value for the distribution of predicted outcomes for structured content records associated with a token produced by the transformation rule, relative to a distribution of predicted outcomes of structured content records without the produced token.

10. (Currently Amended) The method of claim 1, further comprising:

determining a statistical measure of accuracy of the predictive model for a class of candidate transformation rules; and  
selecting a class of transformation rules according to the statistical measure of accuracy.

11. (Currently Amended) The method of claim 1, further comprising:

determining a statistical measure of accuracy of the predictive model for a sequence of candidate transformation rules; and  
selecting a sequence of transformation rules according to the statistical measure of accuracy.

12. (Currently Amended) The method of claim 1, further comprising:

determining a statistical measure of accuracy of the predictive model for each candidate transformation rules in a sequence of candidate transformations rules; and  
selecting a transformation rule from the sequence according to the statistical measure of accuracy.

13. (Currently Amended) The method of claim 1, wherein determining for each candidate transformation rule a statistical measure of accuracy of the predictive model comprises:

determining residuals between the predicted outcomes and actual outcomes for the structured content records associated with tokens produced by the candidate transformation rule.

14. (Original) The method of claim 1, wherein the transformation rules are selected from the group consisting of:

tokenization rules;  
stemming rules;  
case folding rules;  
aliasing rules;  
spelling correction rules;  
phrase generation rules;  
feature generalization rules; and  
translation rules.

15. (Original) The method of claim 1, wherein the predictive model is a supervised learning algorithm.

16. (Currently Amended) The method of claim 1, wherein providing a set of source tokens from the unstructured text content of the customer accounts comprises:

parsing the unstructured text content ~~records~~ using an initial set of transformation rules to produce the set of source tokens; and  
subsequent to the selection of transformation rules, re-parsing the unstructured text content to produce a revised set of source tokens.

17. (Original) The method of claim 1, wherein applying candidate transformation rules to a set of source tokens to selectively produce tokens in response to the transformation rules, comprises:

applying a candidate transformation rule to a source token to produce a token;  
associating the produced token with the source token;  
associating the produced token with the structured content records associated with the source token.

18. (Currently Amended) A method for selecting transformation rules for application to unstructured text content in customer accounts, comprising:

providing a plurality of customer accounts, each customer account comprising:  
a structure content record of financial and personal information associated  
with a customer;  
unstructured text content derived from an interaction with the customer;  
and  
a predicted outcome from a predictive model, wherein the predictive  
model predicts outcomes of events in customer accounts  
providing an index of source tokens from the unstructured text content, each  
source token associated with at least one of the structured content records,

~~each structured content record including a predicted outcome from a predictive model;~~  
 applying candidate transformation rules to the source tokens to selectively produce tokens in response to the transformation rules,  
 associating each token produced by a transformation rule with the structured content records associated with a source token;  
 determining for each transformation rule a statistical measure of the accuracy of the predicted outcomes from the structured content records associated with the tokens produced by the transformation rule; and  
 selecting transformation rules that improve the statistical measure of accuracy of predicted outcomes the predicted model.

19. (Currently Amended) A computer implemented software system for selection of content transformation rules for application to unstructured text content in customer accounts, the system comprising:

a database of customer accounts, each customer account comprising a structure content record of financial and personal information associated with a customer;  
unstructured text content derived from an interaction with the customer;  
and  
a predicted outcome of an event related to the customer; ~~structured content records, each content record including a predicted outcome;~~  
 an index of source tokens derived from the unstructured text content of the customer accounts, each source token associated with at least one of the structured content records;  
 a database of content transformation rules, each transformation rule adapted to produce a token in response to a source token;

a predictive model, adapted to generate the predicted ~~outcome for a~~ outcomes of events related to the customers using the structured content records and tokens derived from the unstructured text content using the content transformation rules; and

a rules selection process, adapted to apply selected transformation rules to the index to produce tokens from the source tokens, and identify transformation rules likely to improve that improve the accuracy of the predictive model.

20. (Original) The system of claim 19, wherein the rules selection process associates each token produced by a transformation rule from a source token with structured content records associated with a source token.

21. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule a number of correct and incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule.

22. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule a distribution of correct and incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule.

23. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by selecting transformation rules that maximize a statistical measure of accuracy of the predictive model.

24. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule:

- a number of correct predicted outcomes from the structured content records associated with a token produced by the transformation rule;
- a number of correct predicted outcomes from the structured content records not associated with the produced token;
- a number of incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule; and
- a number of incorrect predicted outcomes from the structured content records not associated with the produced token.



25. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule an information gain resulting from transformation rule.

26. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule an Odds ratio for correct predicted outcomes in structured content records associated with a token produced by the transformation rule.

27. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule a Chi-square value for the distribution of predicted outcomes for structured content records associated with a token produced by the transformation rule, relative to a distribution of predicted outcomes of structured content records without the produced token.

28. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule a statistical measure of accuracy of the predictive model for a class of candidate transformation rules, and selecting a class of transformation rules according to the measure of accuracy.

29. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule a statistical measure of accuracy of the predictive model for a sequence of candidate transformation rules, and selecting a sequence of transformation rules according to the measure of accuracy.

30. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining for each transformation rule a statistical measure of accuracy of the predictive model

for each candidate transformation rules in a sequence of candidate transformations rules, and selecting a transformation rule from the sequence according to the measure of accuracy.

31. (Currently Amended) The system of claim 19, wherein the rules selection process identifies transformation rules ~~likely to~~ that improve the accuracy of the predictive model by determining residuals between the predicted outcomes and actual outcomes for the structured content records associated with tokens produced by the candidate transformation rule.

32. (Original) The system of claim 19, wherein the transformation rules are selected from the group consisting of:

- tokenization rules;
- stemming rules;
- case folding rules;
- aliasing rules;
- spelling correction rules;
- phrase generation rules;
- feature generalization rules; and
- translation rules.

33. (Original) The system of claim 19, wherein the predictive model is a supervised learning algorithm.

34. (Currently Amended) The system of claim 19, further comprising:

- an indexing process adapted to derive the source tokens for the index from the unstructured text content, and associated each source token with at least one structured content record.

35. (Currently Amended) The system of claim 34, wherein the indexing process is further adapted to:

parse the unstructured text content records using an initial set of transformation rules to produce the index of source tokens ; and  
subsequent to the selection of transformation rules, re-parse the unstructured text content to produce a revised index of source tokens.

36. (Currently Amended) A computer program product, for selecting transformation rules for application to unstructured text content in customer accounts, and storing program instructions on a computer readable medium, the instructions causing a processor to perform the operations comprising:

storing a plurality of customer accounts, each customer account comprising:  
a structure content record of financial and personal information associated  
with a customer;  
unstructured text content derived from an interaction with the customer;  
and  
an actual outcome of an event related to the customer;  
providing a set of source tokens from the unstructured text content of the  
customer accounts, each source token associated with at least one of the  
structured content records, ~~each structured content record including an~~  
~~actual outcome;~~  
applying candidate transformation rules to a set of source tokens to selectively  
produce tokens in response to the transformation rules;  
determining for each candidate transformation rule a statistical measure of  
accuracy of ~~the a~~ predictive model for predicting outcomes of events  
related to the customers based on the actual outcomes of events in the  
customer accounts associated with the produced tokens; and  
selecting transformation rules ~~that are likely~~ to improve the measure of accuracy  
of the predictive model.

37. (Original) The computer program product of claim 36, wherein operations performed by the processor further comprise:

associating each token produced by a transformation rule from a source token with structured content records associated with a source token.

38. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for determining for each candidate transformation rule a statistical measure of accuracy further comprise:

determining a number of correct and incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule.

39. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for determining for each candidate transformation rule a statistical measure of accuracy further comprise:

determining a distribution of correct and incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule.

40. (Original) The computer program product of claim 36, wherein operations performed by the processor for selecting transformation rules further comprise:

selecting transformation rules that maximize the measure of accuracy of the predictive model.

41. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for determining for a candidate transformation rule a statistical measure of accuracy of the predictive model further comprise:

determining a number of correct predicted outcomes from the structured content records associated with a token produced by the transformation rule;

determining a number of correct predicted outcomes from the structured content records not associated with the produced token;  
determining a number of incorrect predicted outcomes from the structured content records associated with a token produced by the transformation rule; and  
determining a number of incorrect predicted outcomes from the structured content records not associated with the produced token.

42. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for determining for each candidate transformation rule a statistical measure of accuracy of the predictive model further comprise:

determining an information gain resulting from transformation rule.

43. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for determining for each candidate transformation rule a statistical measure of accuracy of the predictive model further comprise:

determining an Odds ratio for correct predicted outcomes in structured content records associated with a token produced by the transformation rule.

44. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for determining for each candidate transformation rule a statistical measure of accuracy of the predictive model further comprise:

determining a Chi-square value for the distribution of predicted outcomes for structured content records associated with a token produced by the transformation rule, relative to a distribution of predicted outcomes of structured content records without the produced token.

45. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor further comprise:

determining a statistical measure of accuracy of the predictive model for a class of candidate transformation rules; and

selecting a class of transformation rules according to the measure of accuracy.

46. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor further comprise:

determining a statistical measure of accuracy of the predictive model for a  
sequence of candidate transformation rules; and  
selecting a sequence of transformation rules according to the measure of accuracy.

47. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor further comprise:

determining a statistical measure of accuracy of the predictive model for each  
candidate transformation rules in a sequence of candidate transformations  
rules; and  
selecting a transformation rule from the sequence according to the measure of  
accuracy.

48. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for determining for each candidate transformation rule a statistical measure of accuracy of the predictive model further comprise:

determining residuals between the predicted outcomes and actual outcomes for the  
structured content records associated with tokens produced by the  
candidate transformation rule.

49. (Original) The computer program product of claim 36, wherein the transformation rules are selected from the group consisting of:

tokenization rules;  
stemming rules;  
case folding rules;  
aliasing rules;  
spelling correction rules;

phrase generation rules;  
feature generalization rules; and  
translation rules.

50. (Original) The computer program product of claim 36, wherein the predictive model is a supervised learning algorithm.

51. (Currently Amended) The computer program product of claim 36, wherein operations performed by the processor for providing a set of source tokens from the unstructured text content of the customer accounts further comprise:

parsing the unstructured text content ~~records~~ using an initial set of transformation rules to produce the set of source tokens ; and  
subsequent to the selection of transformation rules, re-parsing the unstructured text content to produce a revised set of source tokens.

52. (Original) The computer program product of claim 36, wherein operations performed by the processor for applying candidate transformation rules to a set of source tokens to selectively produce tokens in response to the transformation rules, further comprise:

applying a candidate transformation rule to a source token to produce a token;  
associating the produced token with the source token;  
associating the produced token with the structured content records associated with the source token.